Multi-class Graph Mumford-Shah Model for Plume Detection using the MBO scheme*

Huiyi Hu[1], Justin Sunu[2], and Andrea L. Bertozzi[1]

[1] Department of Mathematics, University of California, Los Angeles
huiyihu@math.ucla.edu bertozzi@math.ucla.edu
[2] Institute of Mathematical Science, Claremont Graduate University
justinsunu@gmail.com

Abstract. We focus on the multi-class segmentation problem using the piecewise constant Mumford-Shah model in a graph setting. After formulating a graph version of the Mumford-Shah energy, we propose an efficient algorithm called the MBO scheme using threshold dynamics. Theoretical analysis is developed and a Lyapunov functional is proven to decrease as the algorithm proceeds. Furthermore, to reduce the computational cost for large datasets, we incorporate the Nyström extension method which efficiently approximates eigenvectors of the graph Laplacian based on a small portion of the weight matrix. Finally, we implement the proposed method on the problem of chemical plume detection in hyper-spectral video data.

Keywords: graph, segmentation, Mumford-Shah, total variation, MBO, Nyström, hyper-spectral image

1 Introduction

Multi-class segmentation has been studied as an important problem for many years in various areas, such as computer science and machine learning. For imagery data in particular, the Mumford-Shah model [18] is one of the most extensively used model in the past decade. This model approximates the true image by an optimal piecewise smooth function through solving a energy minimization problem. More detailed review of the work on Mumford-Shah model can be found in the references of [4]. A simplified version of Mumford-Shah is the piecewise constant model (also known as the "minimal partition problem"), which is widely used due to its reduced complexity compared to the original one. For a given contour Φ which segments an image region Ω into $\hat{n}$ many disjoint subregions $\Omega = \cup_{r=1}^{\hat{n}} \Omega_r$, the *piecewise constant Mumford-Shah* energy is defined as:

$$E^{\mathrm{MS}}(\Phi, \{c_r\}_{r=1}^{\hat{n}}) = |\Phi| + \lambda \sum_{r=1}^{\hat{n}} \int_{\Omega_r} (u_0 - c_r)^2 \, , \tag{1}$$

* This work was supported by NSF grants DMS-1118971, DMS-1045536, and DMS-1417674, UC Lab Fees Research grant 12-LR-236660, ONR grants N000141210040 and N000141210838, and the W. M. Keck Foundation.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
OCT 2014		**00-00-2014 to 00-00-2014**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Multi-class Graph Mumford-Shah Model for Plume Detection using the MBO scheme	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of California, Los Angeles,Department of Mathematics,Los Angeles,CA,90095**	8. PERFORMING ORGANIZATION REPORT NUMBER **CAM14-79**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We focus on the multi-class segmentation problem using the piecewise constant Mumford-Shah model in a graph setting. After formulating a graph version of the Mumford-Shah energy, we propose an e cient algorithm called the MBO scheme using threshold dynamics. Theoretical analysis is developed and a Lyapunov functional is proven to decrease as the algorithm proceeds. Furthermore, to reduce the computational cost for large datasets, we incorporate the Nystr om extension method which e ciently approximates eigenvectors of the graph Laplacian based on a small portion of the weight matrix. Finally, we implement the proposed method on the problem of chemical plume detection in hyper-spectral video data.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**14**	

where u_0 is the observed image data, $\{c_r\}_{r=1}^{\hat{n}}$ is a set of constant values, and $|\Phi|$ denotes the length of the contour Φ. By minimizing the energy E^{MS} over Φ and $\{c_r\}_{r=1}^{\hat{n}}$, one obtains an optimal function which is constant within each sub-region to approximate u_0, along with a segmentation given by the optimal Φ. In [5], a method of active contours without edges is proposed to solve for the two-class piecewise constant Mumford-Shah model ($\hat{n} = 2$), using a level set method introduced in [19]. The work in [5] is further generalized to a multi-class scenario in [24]. The method developed in [5, 24] is well known as the Chan-Vese model, which is a popular and representative method for image segmentation. The Chan-Vese method has been widely used due to the model's flexibility and the great success it achieves in performance.

In this work, we formulate the piecewise constant MS problem in a graph setting instead of a continuous one, and propose an efficient algorithm to solve it. Recently the authors of [2] introduced a binary semi-supervised segmentation method based on minimizing the Ginzburg-Landau functional on a graph. Inspired by [2], a collection of work has been done on graph-based high-dimensional data clustering problems posed as energy minimization problems, such as semi-supervised methods studied in [14, 11] and an unsupervised network clustering method [13] known as modularity optimization. These methods make use of graph tools [6] and efficient graph algorithms, and our work pursues similar ideas. Note that unlike the Chan-Vese model which uses $log_2(\hat{n})$ many level set functions and binary representations to denote multiple classes, our model uses simplex constrained vectors for class assignments representation (details explained below).

To solve the multi-class piecewise constant MS variational problem in the graph setting, we propose an efficient algorithm using threshold dynamics. This algorithm is a variant of the one presented in the work of Merriman, Bence and Osher (MBO) [16, 17], which was introduced to approximate the motion of an interface by its mean curvature in a continuous space. The idea of the MBO scheme is used on the continuous MS model [8, 21] motivated by level set methods. The authors of [11, 13, 14] implement variants of the MBO scheme applied to segmentation problems in a graph setting. Rigorous proofs of convergence of the original MBO scheme in continuous setting can be found in [1, 9] for the binary case, and [7] for the multi-class case. An analogous discussion in a graph setting is given in [23]. Inspired by the work of [7, 23], we develop a Lyapunov functional for our proposed variant of the MBO algorithm, which approximates the graph MS energy. Theoretical analysis is given to prove that this Lyapunov energy decreases at each iteration of our algorithm, until it converges within finitely many steps.

In order to solve for each iteration of the MBO scheme, one needs to compute the weight matrix of the graph as well as the eigenvectors of the corresponding graph Laplacian. However, the computational cost can become prohibitive for large datasets. To reduce the numerical expenses, we implement the Nyström extension method [10] to approximately compute the eigenvectors, which only requires computing a small portion of the weigh matrix. Thus the proposed

algorithm is efficient even for large datasets, such as the hyper-spectral video data considered in this paper.

The proposed method can be implemented on general high-dimensional data clustering problems. However, in this work the numerical experiment is focused on the detection of chemical plumes in hyper-spectral video data. Detecting harmful gases and chemical plumes has wide applicability, such as in environmental study, defense and national security. However, the diffusive nature of plumes poses challenges and difficulties for the problem. One popular approach is to take advantage of hyper-spectral data, which provides much richer sensing information than ordinary visual images. The hyper-spectral images used in this paper were taken from video sequences captured by long wave infrared (LWIR) spectrometers at a scene where a collection of plume clouds is released. Over 100 spectral channels at each pixel of the scene are recorded, where each channel corresponds to a particular frequency in the spectrum ranging from 7,820 nm to 11,700 nm. The data is provided by the Applied Physics Laboratory at Johns Hopkins University, (see more details in [3]). Prior analysis of this dataset can be found in the works [12, 15, 20, 22]. The authors of [15] implement a semi-supervised graph model using a similar MBO scheme. In the present paper, each pixel is considered as a node in a graph, upon which the proposed unsupervised segmentation algorithm is implemented. Competitive results are achieved as demonstrated below.

The rest of this paper is organized as follows. Section 2 introduces the graph formula for the multi-class piecewise constant Mumford-Shah model and relevant notations. In Secion 3, the Mumford-Shah MBO scheme is presented as well as the theoretical analysis for a Lyapunov functional which is proven to decrease as the algorithm proceeds; techniques such as Nyström method are also introduced for the purpose of numerical efficiency. In Section 4, our algorithm is tested on the hyper-spectral video data for plume detection problem. The results are then presented and discussed.

2 Graph Mumford-Shah Model

Consider an N-node weighted graph (G, E), where $G = \{n_1, n_2, \ldots, n_N\}$ is a node set and $E = \{w_{ij}\}_{i=1}^{N}$ an edge set. Each node n_i corresponds to an agent in a given dataset, (such as a pixel in an image). The quantity w_{ij} represents the similarity between a pair of nodes n_i and n_j. Let $\mathbf{W} = [w_{ij}]$ denote the graph's $N \times N$ *weight matrix*, and in this work we assume $\mathbf{W}$ is symmetric, i.e. $w_{ij} = w_{ji}$. In the case of hyper-spectral data, each node (pixel) n_i is associated with a d-dimensional feature vector (spectral channels). Let $u_0 : G \to \mathbb{R}^d$ denote the raw hyper-spectral data, where $u_0(n_i)$ represents the d-dimensional spectral channels of n_i. We use the following notation:

- The matrix $\mathbf{L} := \mathbf{D} - \mathbf{W}$ is called the (un-normalized) *graph Laplacian* [6], where $\mathbf{D}$ is a diagonal matrix with the i-th entry being $\sum_{j=1}^{N} w_{ij}$. For

$v : G \to \mathbb{R}$, observe that

$$\langle v, \mathbf{L}v \rangle = \frac{1}{2} \sum_{i,j=1}^{N} w_{ij} \left(v(n_i) - v(n_j) \right)^2 . \tag{2}$$

– Graph function spaces for $f = (f_1, f_2, \ldots, f_{\hat{n}}) : G \to \mathbb{R}^{\hat{n}}$:

$$\mathbb{K} := \left\{ f |\ f : G \to [0,1]^{\hat{n}}, \sum_{r=1}^{\hat{n}} f_r(n_i) = 1 \right\} ,$$

which is a compact and convex set.

$$\mathbb{B} := \left\{ f |\ f : G \to \{0,1\}^{\hat{n}}, \sum_{r=1}^{\hat{n}} f_r(n_i) = 1 \right\} \in \mathbb{K} .$$

This simplex constrained vector value taken by $f \in \mathbb{B}$ indicates class assignment, i.e. if $f_r(n_i) = 1$ for some r, then n_i belongs to the r-th class. Thus for each $f \in \mathbb{B}$, it corresponds to a partition of the graph G with at most $\hat{n}$ classes. Let $\langle f, \mathbf{L}f \rangle = \sum_{r=1}^{\hat{n}} \langle f_r, \mathbf{L}f_r \rangle$.

– *Total Variation* (TV) for graph G is given as:

$$|f|_{TV} := \frac{1}{2} \sum_{i,j=1}^{N} w_{ij} \left| f(n_i) - f(n_j) \right| . \tag{3}$$

In this setting, we present a graph version of the multi-class *piecewise constant Mumford-Shah* energy functional:

$$MS(f, \{c_r\}_{r=1}^{\hat{n}}) := \frac{1}{2} |f|_{TV} + \lambda \sum_{r=1}^{\hat{n}} \langle \|u_0 - c_r\|^2, f_r \rangle , \tag{4}$$

where $\{c_r\}_{r=1}^{\hat{n}} \subset \mathbb{R}^d$, $\|u_0 - c_r\|^2$ denotes an $N \times 1$ vector

$$\left(\|u_0(n_1) - c_r\|^2, \ldots, \|u_0(n_N) - c_r\|^2 \right)^T ,$$

and $\langle \|u_0 - c_r\|^2, f_r \rangle = \sum_{i=1}^{N} f_r(n_i) \|u_0(n_i) - c_r\|^2$. Note that when n_i and n_j belong to different classes, we have $|f(n_i) - f(n_j)| = 2$, which leads to the coefficient in front of the term $\frac{1}{2} |f|_{TV}$.

To see the connection between (4) and (1), one first observes that f_r is the characteristic function of the r-th class, and thus $\langle \|u_0 - c_r\|^2, f_r \rangle$ is analogous to the term $\int_{\Omega_r} (u_0 - c_r)^2$ in (1). Furthermore, the total variation of the characteristic function of a region gives the length of its boundary contour, and therefore $|f|_{TV}$ is the graph analogy of $|\Phi|$.

In order to find a segmentation for G, we propose to solve the following minimization problem:

$$\min_{f \in \mathbb{B}, \{c_r\}_{r=1}^{\hat{n}} \subset \mathbb{R}^d} MS(f, \{c_r\}_{r=1}^{\hat{n}}) . \tag{5}$$

The resulting minimizer f yields a partition of G.

One can observe that the optimal solution of (5) must satisfy:

$$c_r = \frac{\langle u_0, f_r \rangle}{\sum_{i=1}^{N} f_r(n_i)}\,, \tag{6}$$

if the r-th class is non-empty.

Note that for the minimization problem given in (5), it is essentially equivalent to the K-means method when λ goes to $+\infty$. When $\lambda \to 0$, the minimizer approaches a constant.

3 Mumford-Shah MBO and Lyapunov functional

The authors of [16, 17] introduced an efficient algorithm (known as the MBO scheme) to approximate the motion by mean curvature of an interface in a continuous space. The general procedure of the MBO scheme alternates between solving a linear heat equation and thresholding. One interpretation of the scheme is that it replaces the non-linear term of the Allen-Cahn equation with thresholding [8]. In this section we propose a variant of the original *MBO* scheme to approximately find the minimizer of the energy $MS(f, \{c_r\}_{r=1}^{\hat{n}})$ presented in (4). Inspired by the work of [7, 23], we write out a Lyapunov functional $Y_\tau(f)$ for our algorithm and prove that it decreases at each iteration of the MBO scheme.

3.1 Mumford-Shah MBO scheme

We first introduce a "diffuse operator" $\Gamma_\tau = e^{-\tau \mathbf{L}}$, where $\mathbf{L}$ is the graph Laplacian defined above and τ is a time step size. The operator Γ_τ is analogous to the diffuse operator $e^{-\tau \Delta}$ of the heat equation in PDE (continuous space). It satisfies the following properties.

Proposition 1. *Firstly, Γ_τ is strictly positive definite, i.e. $\langle f, \Gamma_\tau f \rangle > 0$ for any $f \in \mathbb{K},\ f \neq 0$. Secondly, Γ_τ conserves the mass, i.e. $\langle 1, \Gamma_\tau f \rangle = \langle 1, f \rangle$. At last, the quantity $\frac{1}{2\tau}\langle 1 - f, \Gamma_\tau f \rangle$ approximates $\frac{1}{2}\,|f|_{TV}$, for any $f \in \mathbb{B}$.*

Proof. Taylor expansion gives

$$e^{-\tau \mathbf{L}} = I - \tau \mathbf{L} + \frac{\tau^2}{2!}\mathbf{L}^2 - \frac{\tau^3}{3!}\mathbf{L}^3 + \dots$$

Suppose v is an eigenvector of $\mathbf{L}$ associated with the eigenvalue ξ. One then has $\Gamma_\tau v = e^{-\tau\xi} v \ \Rightarrow \langle v, \Gamma_\tau v \rangle = e^{-\tau\xi}\langle v, v \rangle > 0$. Let the eigen-decomposition (with respect to $\mathbf{L}$) for a non-zero $f : G \to \mathbb{R}$ to be $f = \sum_{i=1}^{N} a_i \phi_i$, where $\{\phi_i\}_{i=1}^{N}$ is a set of orthogonal eigenvectors of $\mathbf{L}$ (note that $\mathbf{L}$ is positive definite). Because Γ_τ is a linear operator, one therefore has $\langle f, \Gamma_\tau f \rangle = \sum_{i=1}^{N} a_i^2 \langle \phi_i, \Gamma_\tau \phi_i \rangle > 0$.

For the second property, $\mathbf{L}\mathbf{1} = 0 \Rightarrow \langle \mathbf{1}, \mathbf{L}^k f \rangle = 0$, where $\mathbf{1}$ is an N-dimensional vector with one at each entry. Therefore, the Taylor expansion of Γ_τ gives $\langle \mathbf{1}, \Gamma_\tau f \rangle = \langle \mathbf{1}, f \rangle$.

At last, $\Gamma_\tau \simeq I - \tau\mathbf{L} \Rightarrow \frac{1}{2\tau}\langle 1-f, \Gamma_\tau f\rangle \simeq \frac{1}{2\tau}\langle 1-f, f\rangle - \frac{1}{2}\langle 1, \mathbf{L}f\rangle + \frac{1}{2}\langle f, \mathbf{L}f\rangle$. Particularly when $f \in \mathbb{B}$, we have $\frac{1}{2\tau}\langle 1-f, f\rangle = \frac{1}{2}\langle 1, \mathbf{L}f\rangle = 0$ and $\frac{1}{2}\langle f, \mathbf{L}f\rangle = \frac{1}{2}|f|_{TV}$. Hence $\frac{1}{2\tau}\langle 1-f, \Gamma_\tau f\rangle$ approximates $\frac{1}{2}|f|_{TV}$ in $\mathbb{B}$.

□

Note that the operator $(I+\tau\mathbf{L})^{-1}$ also satisfies the above three properties, and can serve the same purpose as $e^{-\tau\mathbf{L}}$, as far as this paper concerns.

The proposed *Mumford-Shah MBO* scheme for the minimization problem (5) consists of alternating between the following three steps:

For a given $f^k \in \mathbb{B}$ at the k-th iteration and $c_r^k = \frac{\langle u_0, f_r^k\rangle}{\langle \mathbf{1}, f_r^k\rangle}$,

1. Compute
$$\hat{f} = \Gamma_\tau f^k - \tau\lambda\left(\|u_0 - c_1^k\|^2, \|u_0 - c_2^k\|^2, \ldots, \|u_0 - c_{\hat{n}}^k\|^2\right), \tag{7}$$
2. (Thresholding)
$$f^{k+1}(n_i) = \boldsymbol{e}_r, \quad r = \text{argmax}_c \hat{f}_c(n_i)$$
for all $i \in \{1, 2, \ldots, N\}$, where $\boldsymbol{e}_r$ is the r-th standard basis in $\mathbb{R}^{\hat{n}}$, i.e. $f_r^{k+1}(n_i) = 1$ and $f^{k+1} \in \mathbb{B}$.
3. (Update c)
$$c_r^{k+1} = \frac{\langle u_0, f_r^{k+1}\rangle}{\langle \mathbf{1}, f_r^{k+1}\rangle}.$$

3.2 A Lyapunov functional

We introduce a Lyapunov functional Y_τ for the Mumford-Shah MBO scheme:

$$Y_\tau(f) := \frac{1}{2\tau}\langle 1-f, \Gamma_\tau f\rangle + \lambda\sum_{r=1}^{\hat{n}}\langle \|u_0 - c_r\|^2, f_r\rangle, \quad \text{subject to } c_r = \frac{\langle u_0, f_r\rangle}{\langle \mathbf{1}, f_r\rangle}. \tag{8}$$

According to the third property of Γ_τ in Proposition 1, energy $Y_\tau(f)$ approximates $MS(f, \{c_r\}_{r=1}^{\hat{n}})$ for $f \in \mathbb{B}$ and $c_r = \frac{\langle u_0, f_r\rangle}{\langle \mathbf{1}, f_r\rangle}$. A similar functional for the graph total variation is shown and discussed in [23].

Pursuing similar ideas as in [7, 23], we present the following analysis which consequently shows that the Mumford-Shah MBO scheme (with time step τ) decreases Γ_τ and converges to a stationary state within a finite number of iterations.

First define

$$G_\tau(f, c) := \frac{1}{2\tau}\langle 1-f, \Gamma_\tau f\rangle + \lambda\sum_{r=1}^{\hat{n}}\langle \|u_0 - c_r\|^2, f_r\rangle. \tag{9}$$

Proposition 2. *The functional $G_\tau(\cdot, c)$ is strictly concave on $\mathbb{K}$, for any fixed $\{c_r\}_{r=1}^{\hat{n}} \in \mathbb{R}^d$.*

Proof. Take $f, g \in \mathbb{K}$, $\alpha \in (0,1)$. We have $(1-\alpha)f + \alpha g \in \mathbb{K}$, because $\mathbb{K}$ is a convex set.

$$\begin{aligned} & G_\tau\left((1-\alpha)f + \alpha g, c\right) - (1-\alpha)G_\tau(f,c) - \alpha G_\tau(g,c) \\ = & \frac{1}{2\tau}\alpha(1-\alpha)\langle f-g, \Gamma_\tau(f-g)\rangle \geq 0\,. \end{aligned} \tag{10}$$

Equality only holds when $f = g$. Therefore, $G_\tau(\cdot, c)$ is strictly concave on $\mathbb{K}$. □

Aside from the concavity of G_τ, we observe that the first order variation of $G_\tau(\cdot, c)$ is given as

$$\frac{\delta}{\delta f}G_\tau(f,c) = \frac{1}{2\tau}(1 - 2\Gamma_\tau f) + \lambda\left(\|u_0 - c_1\|^2, \|u_0 - c_2\|^2, \ldots\right)\,.$$

Note that since $\langle \frac{\delta}{\delta f}G_\tau(f^k, c^k), f\rangle$ is linear, the Step 2 (thresholding) in the Mumford-Shah MBO scheme is equivalent to

$$f^{k+1} := \mathrm{argmin}_{f\in K}\langle \frac{\delta}{\delta f}G_\tau(f^k, c^k), f\rangle\,.$$

Theorem 1. *In the Mumford-Shah MBO scheme, the Lyapunov functional $Y_\tau(f^{k+1})$ at the $(k+1)$-th iteration is no greater than $Y_\tau(f^k)$. Equality only holds when $f^k = f^{k+1}$. Therefore, the scheme achieves a stationary point in $\mathbb{B}$ within a finite number of iterations.*

Proof.

$$f^{k+1} := \mathrm{argmin}_{f\in K}\langle \frac{\delta}{\delta f}G_\tau(f^k, c^k), f\rangle \tag{11}$$

$\Rightarrow f^{k+1} \in \mathbb{B}$ (due to linearity) and

$$\begin{aligned} 0 &\geq \langle \frac{\delta}{\delta f}G_\tau(f^k, c^k), f^{k+1} - f^k\rangle \\ &\geq G_\tau(f^{k+1}, c^k) - G_\tau(f^k, c^k) \qquad \text{(concavity)} \end{aligned} \tag{12}$$

$\Rightarrow G_\tau(f^{k+1}, c^k) \leq G_\tau(f^k, c^k) = Y_\tau(f^k)$. Observe that $c_r^{k+1} = \frac{\langle f_r^{k+1}, u_0\rangle}{\langle f_r^{k+1}, 1\rangle}$ is the minimizer of

$$\mathrm{argmin}_{\{c_r\}_{r=1}^{\hat{n}}\in\mathbb{R}^d} G_\tau(f^{k+1}, c)$$

$\Rightarrow G_\tau(f^{k+1}, c^{k+1}) \leq G_\tau(f^{k+1}, c^k) \leq Y_\tau(f^k)$.
$\Rightarrow Y_\tau(f^{k+1}) \leq Y_\tau(f^k)$. Therefore the Lyapunov functional Y_τ is decreasing on the iterations of the Mumford-Shah MBO scheme, unless $f^{k+1} = f^k$. Since $\mathbb{B}$ is a finite set, a stationary point can be achieved in a finite number of iterations. □

Minimizing the Lyapunov energy Γ_τ is an approximation of the minimization problem in (5), and the proposed MBO scheme is proven to decrease Γ_τ. Therefore, we expect the Mumford-Shah MBO scheme to approximately solve (5). In Section 3.3 and Section 3.4, we introduce techniques for computing the MBO iterations efficiently.

3.3 Eigen-space approximation

To solve for (7) in Step 1 of the Mumford-Shah MBO scheme, one needs to compute the operator Γ_τ, which can be difficult especially for large datasets. For the purpose of efficiency, we numerically solve for (7) by using a small number of the leading eigenvectors of $\mathbf{L}$ (which correspond to the smallest eigenvalues), and project f^k onto the eigen-space spanned from the eigenvectors. By approximating the operator $\mathbf{L}$ with the leading eigenvectors, one can compute (7) efficiently. We use this approximation because in graph clustering methods, researchers have been using a small portion of the leading eigenvectors of a graph Laplacian to extract structural information of the graph.

Let $\{\phi_m\}_{m=1}^M$ denote the first M (orthogonal) leading eigenvectors of $\mathbf{L}$, and $\{\xi_m\}_{m=1}^M$ the corresponding eigenvalues. Assume $f^k = \sum_{m=1}^M \phi_m a^m$, where a^m is a $1\times\hat{n}$ vector, with the r-th entry $a_r^m = \langle f_r^k, \phi_m\rangle$. Thus $\hat{f}$ can be approximately computed as:

$$\hat{f} = \sum_{m=1}^{M} e^{-\tau\xi_m}\phi_m a^m - \tau\lambda\left(\|u_0 - c_1^k\|^2, \|u_0 - c_2^k\|^2, \ldots, \|u_0 - c_{\hat{n}}^k\|^2\right). \quad (13)$$

The Mumford-Shah MBO algorithm with the above eigen-space approximation is summarized in Algorithm 1. After the eigenvectors are obtained, each iteration of the MBO scheme is of time complexity $O(N)$. Empirically, the algorithm converges after a small number of iterations. Note that the iterations stop when a *purity* score between the partitions from two consecutive iterations is greater than 99.9%. The purity score, as used in [13], measures how "similar" two partitions are. Intuitively, it can be viewed as the fraction of nodes of one partition that have been assigned to the correct class with respect to the other partition.

Algorithm 1 Mumford-Shah MBO algorithm

Input: f^0, u_0, $\{(\phi_m, \xi_m\}_{m=1}^M$, τ, λ, $\hat{n}$, $k = 0$.
while (purity$(f^k, f^{k+1}) < 99.9\%$) **do**
- $c_r = \frac{\langle u_0, f_r^k\rangle}{\sum_{i=1}^N f_r^k(n_i)}$.
- $a_r^m = \langle f_r^k, \phi_m\rangle$.
- $\hat{f} = \sum_{m=1}^M e^{-\tau\xi_m}\phi_m a^m - \tau\lambda\left(\|u_0 - c_1\|^2, \|u_0 - c_2\|^2, \ldots, \|u_0 - c_{\hat{n}}\|^2\right)$.
- $f^{k+1}(n_i) = \mathbf{e}_r$, where $r = \text{argmax}_c \hat{f}_c(n_i)$.
- $k \leftarrow k + 1$.

end while

3.4 Nyström method

The Nyström extension [10] is a matrix completion method which has been used to efficiently compute a small portion of the eigenvectors of the graph Laplacian for segmentation problems [2, 14, 11]. In our proposed scheme, leading

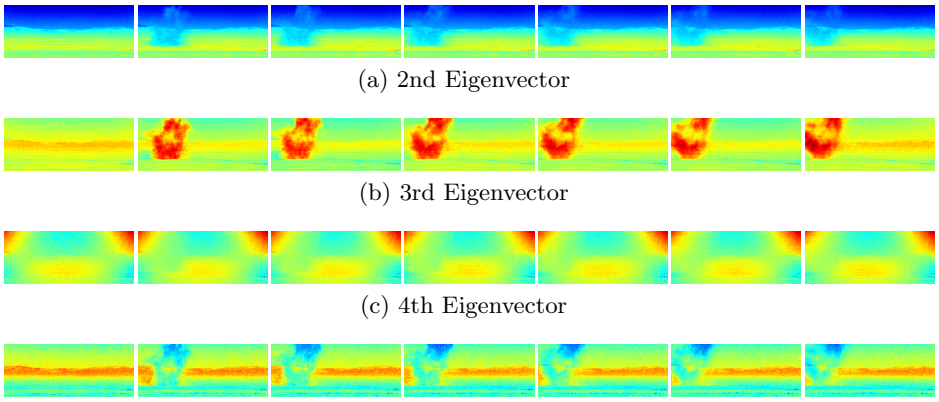
eigenvectors of $\mathbf{L}$ are required, which can require massive computational time and memory. For large graphs such as the ones induced from images, the explicit form of the weight matrix $\mathbf{W}$ and therefore $\mathbf{L}$ is difficult to obtain ($O(N^2)$ time complexity). Hence, we expect to use the Nyström method to approximately compute the eigenvectors for our algorithm.

Basically, the Nyström method randomly samples a very small number (M) of rows of $\mathbf{W}$. Based on matrix completion and properties of eigenvectors, it approximately obtains M eigenvectors of the symmetric normalized graph Laplacian $\mathbf{L}_s = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}}$ without computing the whole weight matrix. Detailed descriptions of the Nyström method can be found in [2, 14].

Note that our previous analysis only applies to $\mathbf{L}$ rather than $\mathbf{L}_s$, and the Nyström method can not be trivially formularized for $\mathbf{L}$. Therefore this question remains to be studied. However, the normalized Laplacian $\mathbf{L}_s$ has many similar features compared to $\mathbf{L}$, and it has been used in place of $\mathbf{L}$ in many segmentation problems. In the numerical results shown below, the eigenvectors of $\mathbf{L}_s$ computed via Nyström perform well empirically.

One can also implement other efficient methods to compute the eigenvectors for the Mumford-Shah MBO algorithm.

4 Numerical Results



(a) 2nd Eigenvector

(b) 3rd Eigenvector

(c) 4th Eigenvector

(d) 5th Eigenvector

Fig. 1: The leading eigenvectors of the normalized graph Laplacian computed via the Nyström method.

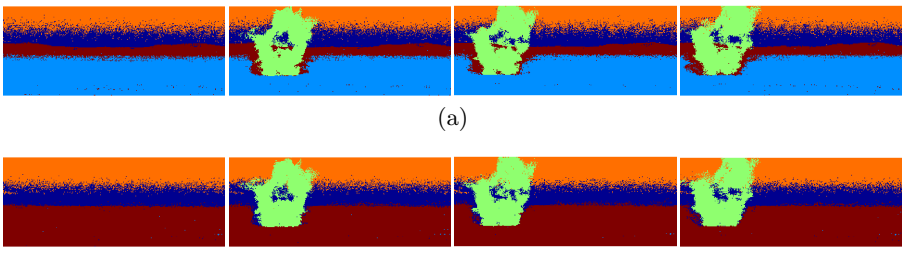
The hyper-spectral images tested in this work are taken from the video recording of the release of chemical plumes at the Dugway Proving Ground, captured by long wave infrared (LWIR) spectrometers. The data is provided by the Applied Physics Laboratory at Johns Hopkins University. A detailed description of this dataset can be found in [3]. We take seven frames from a plume video

sequence in which each frame is composed of 128×320 pixels. We use a background frame and the frames numbered 72 through 77 containing the plume. Each pixel has 129 spectral channels corresponding to a particular frequency in the EM spectrum ranging from 7,820 nm to 11,700 nm. Thus, the graph we construct from these seven frames is of size $7 \times 128 \times 320$ with each node n_i corresponding to a pixel with a 129-dimensional spectral signature v_i. The metric for computing the weight matrix is given as:

$$w_{ij} = 1 - \frac{\langle v_i, v_j \rangle}{\|v_i\| \|v_j\|},$$

which is an approximation of the spectral angle $\cos(v_i, v_j)$.

The goal is to segment the image and identify the "plume cloud" from the background components (sky, mountain,grass), without any ground truth. As described in the previous section, $M = 100$ eigenvectors of the normalized graph Laplacian ($\mathbf{L}_s$) are computed via the Nyström method. The computational time using Nyström is less than a minute on a 2.8GHz machine with Intel Core 2Duo. The visualization of the first five eigenvectors (associated with the smallest eigenvalues) are given in Figure 1 for the first four frames, (the first eigenvector is not shown because it is close to a constant vector).



(a)

(b)

Fig. 2: The segmentation results obtained by the Mumford-Shah MBO scheme, on a background frame plus the frames 72-77. Shown in (a) and (b) are segmentation outcomes obtained with different initializations. The visualization of the segmentations only includes the first four frames.

We implement the Mumford-Shah MBO scheme using the eigenvectors on this seven frames of plume images, with $\tau = 0.15$, $\lambda = 150$ and $\hat{n} = 5$. The test is run for 20 times with different uniformly random initialization, and the segmentation results are shown in Figure 2. Note that depending on the initialization, the algorithm can converge to different local minimum, which is common for most non-convex variational methods. The result in (a) occurred five times among the 20 runs, and (b) for twice. The outcomes of other runs merge either the upper or the lower part of the plume with the background. The segmentation outcome shown in (a) gives higher energy than that in (b). Among the 20

runs, the lowest energy is achieved by a segmentation similar to (a), but with the lower part of the plume merged with the background. It may suggest that the global minimum of the proposed energy does not necessarily give a desired segmentation.

Notice that in Figure 2 (b), even though there actually exist five classes, only four major classes can be perceived, while the other one contains only a very small amount of pixels. By allowing $\hat{n} = 5$ instead of 4, it helps to reduce the influence of a few abnormal pixels. The computational time for each iteration is about 2-3 seconds on a 1.7GHz machine with Intel Core i5. The number of iterations is around 20-40.

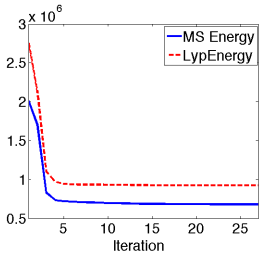
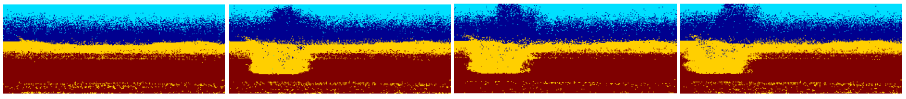


Fig. 3: Energy $MS(f)$ (blue, solid line) and $Y_\tau(f)$ (red, dash line) at each iteration from the same test as shown in Figure 2 (a).

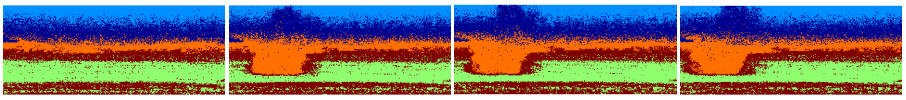
Figure 3 demonstrates a plot of the $MS(f)$ and $Y_\tau(f)$ energies at each iteration from the same test as the one shown in Figure 2 (a). The Lyapunov energy $Y_\tau(f)$ (red, dash line) is non-increasing, as proven in Theorem 1. Note that all the energies are computed approximately using eigenvectors.

As a comparison, the segmentation results using K-means and spectral clustering are shown in Figure 4. The K-means method is performed directly on the raw image data ($7 \times 128 \times 320$ by 129). As shown in (a) and (b), the results obtained by K-means fail to capture the plume; the segmentations on the background are also very fuzzy. For the spectral clustering method, a 4-way (or 5-way) K-means is implemented on the four (or five) leading eigenvectors of the normalized graph Laplacian (computed via Nyström). As shown in (c) and (d), the resulting segmentations divide the sky region into two undesirable components. Unlike the segmentation in Figure 2 (a) where the mountain component (red, the third in the background) has a well defined outline, the spectral clustering results do not provide clear boundaries. Our approach performs better than other unsupervised clustering results on this dataset [12, 20].

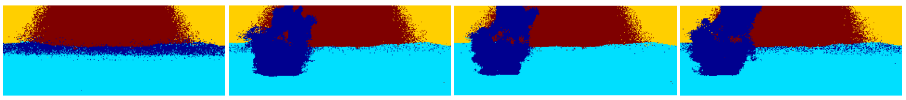
Another example of the plume data is show in Figure 5, where the 67th to 72nd frames (instead of the 72nd to 77th) are taken along with the background frame as the test data. The test is run 20 times using different uniformly random initialization, where $\tau = 0.15$, $\lambda = 150$ and $\hat{n} = 5$. The result in Figure 5 (a) occurred 11 times among the 20 runs, and (b) for 5 times. The outcomes from the other 4 runs segment the background into three components as in (a), but



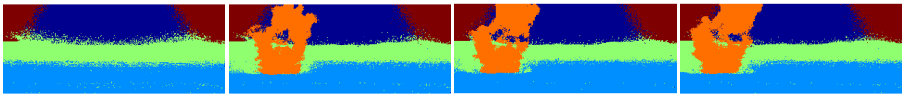
(a) 4-way K-means



(b) 5-way K-means

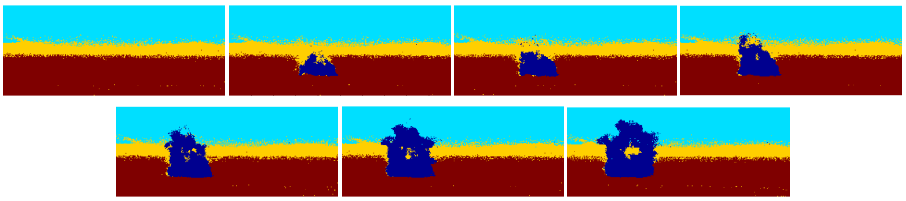


(c) Spectral Clustering with 4-way K-means

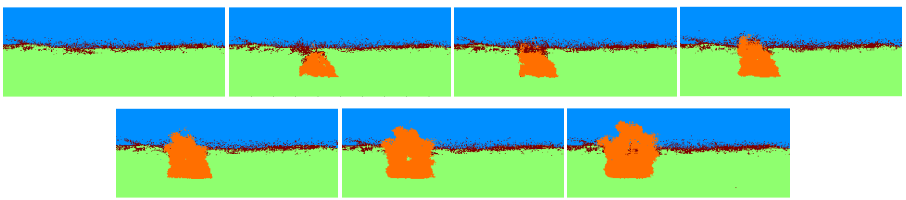


(d) Spectral Clustering with 5-way K-means

Fig. 4: K-means and spectral clustering segmentation results. The visualization of the segmentations only includes the first four frames.



(a)



(b)

Fig. 5: The segmentation results obtained by the Mumford-Shah MBO scheme, on a background frame plus the frames 67-72. Shown in (a) and (b) are segmentation outcomes obtained with different initializations.

merge the plume with the center component. The segmentation result shown in (a) gives the lowest energy among all the outcomes. The visualization includes all seven frames since the plume is small in the first several frames.

5 Conclusion

In this paper we present a graph framework for the multi-class piecewise constant Mumford-Shah model using a simplex constrained representation. Based on the graph model, we propose an efficient threshold dynamics algorithm, the Mumford-Shah MBO scheme for solving the minimization problem. Theoretical analysis is developed to show that the MBO iteration decreases a Lyapunov energy that approximates the MS functional. Furthermore, in order to reduce the computational cost for large datasets, we incorporate the Nyström extension method to approximately compute a small portion of the eigenvectors of the normalized graph Laplacian, which does not require computing the whole weight matrix of the graph. After obtaining the eigenvectors, each iteration of the Mumford-Shah MBO scheme is of time complexity $O(n)$. The number of iterations for convergence is small empirically.

The proposed method can be applied to general high-dimensional data segmentation problems. In this work we focus on the segmentation of hyper-spectral video data. Numerical experiments are performed on a collection of hyper-spectral images taken from a video for plume detection; using our proposed method, competitive results are achieved. However, there are still open questions to be answered. For example, the Nyström method can only compute eigenvectors for the normalized Laplacian, while the theoretical analysis for the Lyapunov functional only applies to the un-normalized graph Laplacian. This issue remains to be studied. Note that the graph constructed in this paper does not include the spacial information of the pixels, but only the spectral information. One can certainly build a graph incorporating the location of each pixel as well, to generate a non-local means graph as discussed in [2].

Acknowledgments. We thank Dr. Luminita A. Vese for useful comments.

References

1. Barles, G., Georgelin, C.: A Simple Proof of Convergence for an Approximation Scheme for Computing Motions by Mean Curvature. SIAM J. Numer. Anal. 32(2), 484–500 (1995)
2. Bertozzi, A. L., Flenner, A.: Diffuse Interface Models on Graphs for Classification of High Dimensional Data, Multiscale Modeling Sim. 10 (3), 1090–1118 (2012)
3. Broadwater, J. B., Limsui, D., Carr, A. K.: A Primer for Chemical Plume Detection using LWIR Sensors. Tech. Rep., National Security Technology Department, (2011)
4. Cai, X., Chan, R., Zeng, T.: A Two-Stage Image Segmentation Method Using a Convex Variant of the Mumford-Shah Model and Thresholding. SIAM J. Imaging Sci. 6(1), 368–390 (2013)

5. Chan, T., Vese, L. A.: Active Contours without Edges, IEEE Trans. Image Process. 10, 266–277 (2001)
6. Chung, F. R. K.: Spectral Graph Theory, CBMS Reg. Conf. Ser. Math. 92, AMS, (1997)
7. Esedoglu, S., Otto, F.: Threshold Dynamics for Networks with Arbitrary Surface Tensions. Submitted, (2013)
8. Esedoglu, S., Tsai, Y. R.: Threshold Dynamics for the Piecewise Constant Mumford-Shah Functional, J. Comput. Phys. 211, 367–384 (2006)
9. Evans, L. C.: Convergence of an Algorithm for Mean Curvature Motion. Indiana Univ. Math. J. 42, 553–557 (1993)
10. Fowlkes, C., Belongie, S., Chung, F., Malik, J.: Spectral Grouping using the Nystrom Method. IEEE Trans. Pattern Anal. Mach. Intell. 26(2), 214-225 (2004)
11. Garcia-Cardona, C., Merkurjev, E., Bertozzi, A. L., Flenner, A., Percus, A.: Fast Multiclass Segmentation using Diffuse Interface Methods on Graphs. IEEE Trans. Pattern Anal. Mach. Intell. (2014)
12. Gerhart T., Sunu J., Lieu L., Merkurjev E., Chang J.-M., Gilles J., Bertozzi A. L.: Detection and Tracking of Gas Plumes in LWIR Hyperspetral Video Sequence Data. In: SPIE Conference on Defense, Security, and Sensing, (2013)
13. Hu, H., Laurent, T., Porter, M. A., Bertozzi, A. L.: A Method Based on Total Variation for Network Modularity Optimization using the MBO Scheme. SIAM J. Appl. Math. 73(6), 2224–2246 (2013)
14. Merkurjev, E., Kostic, T., Bertozzi, A. L.: An MBO Scheme on Graphs for Segmentation and Image Processing. SIAM J. Imaging Sci. 6, 1903–1930 (2013)
15. Merkurjev, E., Sunu, J., Bertozzi, A. L.: Graph MBO Method for Multiclass Segmentation of Hyper Spectral Stand-off Detection Video. In: Proceedings of the International Conference on Image Processing, accepted, (2014)
16. Merriman, B., Bence, J. K., Osher, S. J.: Diffusion Generated Motion by Mean Curvature. In: Proceedings of the Computational Cristal Growers Workshop, pp. 73–83. AMS, Providence (1992)
17. Merriman, B., Bence, J. K., Osher, S. J.: Motion of Multiple Junctions: A Level Set Approach. J. Comput. Phys. 112, 334–363 (1994)
18. Mumford, D., Shah, J.: Optimal Approximation by Piecewise Smooth Functions and Associated Variational Problems. Comm. Pure Appl. Math. 42, 577–685 (1989)
19. Osher, S., Sethian, J. A.: Fronts Propagating with Curvature Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations. J. Comput. Phys. 79(1), 12–49 (1988)
20. Sunu, J., Chang, J.-M., Bertozzi, A. L.: Simultaneous Spectral Analysis of Multiple Video Sequence Data for LWIR Gas Plumes. In: SPIE Conference on Defense, Security, and Sensing, (2014)
21. Tai, X. C., Christiansen, O., Lin, P., Skjælaaen, I.: Image Segmentation using Some Piecewise Constant Level Set Methods with MBO Type of Projection. In: International Journal of Computer Vision, 73(1), 61–76 (2007)
22. Tochon, G., Chanussot, J., Gilles, J., Dalla Mura, M., Chang, J.-M., Bertozzi, A. L.: Gas Plume Detection and Tracking in Hyperspectral Video Sequences using Binary Partition Trees, preprint, (2014)
23. van Gennip, Y., Guillen, N., Osting, B., Bertozzi, A. L.: Mean Curvature, Threshold Dynamics, and Phase Field Theory on Finite Graphs. Milan J. Math. 82(1), 3–65 (2014)
24. Vese, L. A., Chan, T. F.: A New Multiphase Level Set Framework for Image Segmentation via the Mumford and Shahs model. In: International Journal of Computer Vision 50, 271–293 (2002)